New Methods for Representing and Interacting with Qualitative Geographic Information

Contract #: W9132V-11-P-0010

Contract Period: May 23, 2011 – October 31, 2012

Principal Investigators:

Dr. Alan M. MacEachren, GeoVISTA Center, Penn State University

Dr. Anthony Robinson, GeoVISTA Center, Penn State University

Report on Component 1: Current and Future Methods for Representing and Interacting with Qualitative Geographic Information

Alexander Savelyev, Anthony C. Robinson, and Alan M. MacEachren

<savelyev, arobinson, maceachren>@psu.edu

GeoVISTA Center, Department of Geography, The Pennsylvania State University

Submitted, Oct. 26, 2011

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* **26-10-2011**	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* May 23, 2011 – October 26, 2011

4. TITLE AND SUBTITLE

Report on Component 1: Current and Future Methods for Representing and Interacting with Qualitative Geographic Information

5a. CONTRACT NUMBER
W9132V-11-P-0010

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Alexander Savelyev, Anthony C. Robinson, and Alan M. MacEachren

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

PENNSYLVANIA STATE UNIVERSITY , THE
408 OLD MAIN
UNIVERSITY PARK PA 16802-1505

8. PERFORMING ORGANIZATION REPORT NUMBER

U436430 BAA #4814--Task 1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

US Army Engineer Research and Development Center (ERDC)
Topographic Engineering Center (TEC)
7701 Telegraph Road
Alexandria, VA 22135-3864

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; Distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Geographically-grounded text information is an increasingly common data type that has the potential to increase our ability to understand place-based activities and processes dramatically if methods can be developed to extract, process, and represent that information as well as to connect the information with more traditional geographic data organized within GIS and related technologies. A variety of approaches exist for visual exploration and analysis of text media, and this report highlights and categorizes known approaches towards handling text information in information visualization and geographic information technologies. In addition, we describe the most common techniques for interacting with textual data and its derivatives in geographic and non-geographic visualization systems. Finally, we propose several graphical methods for using text itself to represent different dimensions of geographic information. These methods, as well as others we review from previous work, help elaborate a path forward for future geographic information technologies that can more effectively leverage geographically-grounded text.

15. SUBJECT TERMS

geovisualization, visual analytics, social media, microblogs, cartography, qualitative geographic information

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	25	**19b. TELEPHONE NUMBER** *(include area code)*
SAR	SAR	SAR			

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Component 1: Current and Future Methods for Representing and Interacting with Qualitative Geographic Information

Alexander Savelyev, Anthony C. Robinson, and Alan M. MacEachren
GeoVISTA Center, Department of Geography, The Pennsylvania State University
<savelyev, arobinson, maceachren>@psu.edu

Abstract
Geographically-grounded text information is an increasingly common data type that has the potential to increase our ability to understand place-based activities and processes dramatically if methods can be developed to extract, process, and represent that information as well as to connect the information with more traditional geographic data organized within GIS and related technologies. A variety of approaches exist for visual exploration and analysis of text media, and this report highlights and categorizes known approaches towards handling text information in information visualization and geographic information technologies. In addition, we describe the most common techniques for interacting with textual data and its derivatives in geographic and non-geographic visualization systems. Finally, we propose several graphical methods for using text itself to represent different dimensions of geographic information. These methods, as well as others we review from previous work, help elaborate a path forward for future geographic information technologies that can more effectively leverage geographically-grounded text.

1. Introduction – visual analytics for geospatial information
We begin our review with a broad overview of relevant research domains. A significant proportion of work surveyed for this literature review belongs to the domain of *visual text analytics*, a subclass of a broader field of visual analytics. The latter has been described as *the science of analytical reasoning facilitated by interactive visual interfaces* (Thomas and Cook 2005). Visual text analytics, in turn, applies a combination of analytical tools and interactive visual interfaces to enable reasoning about large collections of textual information (Risch et al. 2008). *Geovisual analytics*, the approach that is the focus of our current research, is similar to information visualization and visual text analytics in that it makes use of interactive interfaces to explore and solve ill-structured problems. Its unique aspect, however, is the explicit treatment of geographic space as one of the key dimensions (MacEachren and Kraak 1997, Andrienko et al. 2007).

This report continues with a broad review of existing visualization approaches, including both representation methods as well as interaction types that have had demonstrated utility in tasks that require the visualization of textual media that may originate from a variety of sources and may include a focus on the structure of documents, their relationships, and their associated entities. We follow this review with a proposed set of new ideas that take advantage of visual variables we can apply to text in order to indicate important aspects of geographic information. Finally, we conclude with our recommendations for future research and development in support of the long-term goal to supply analysts with more efficient and effective means for visual analysis of qualitative geospatial information.

2. Current Approaches for Representing and Interacting with Text

Geovisual analytics, visual analytics of text and visual analytics in general are closely related and share a large number of common aims. Two large themes, namely the investigation of *thematic* and *temporal* characteristics of datasets, can be traced through all three fields. Investigation of and reasoning about geographic space as an explicit dimension is specific to geovisual analytics, as mentioned above. These three themes, along with a list of interface design considerations relevant for the exploration of textual data, are presented in the following sections.

2.1 Thematic characteristics of the dataset

Some of the biggest challenges in textual data visualization result from the thematic richness of the underlying dataset. Although some domains such as network visualization enjoy a certain degree of structure (information related to document title, publication date, authorship and the like is often available), the majority of problems correspond with so-called “freeform” text. One of the simplest and most well-known approaches towards providing a summary of a text document is a tag cloud (Sinclair and Cardew-Hall 2008), a weighted collection of key terms presented to the user in a graphical fashion. Despite the fact that tag clouds are now ubiquitous (Cidell 2010, Lee et al. 2010, Viégas, Wattenberg and Feinberg 2009, Wood et al. 2007), there is considerable argument about whether they are sufficient for any kind of analytical work (Sinclair and Cardew-Hall 2008). Tag clouds aside, two prominent directions in the visual analysis of the freeform text can be identified, namely; spatialization, and visualization of document structure.

2.1.1 Spatialization

The goal of information spatialization is to reduce the number of dimensions in the original dataset to either 2 or 3, making it possible to explore the collection of documents as a two- or three-dimensional landscape, respectively. Relative proximity of documents in the resulting landscape reflects their semantic similarity (Wise, 1999). Spatialization has been a highly productive area of research in the last two decades. Some of the most prominent examples of spatialization systems include SPIRE (Wise 1999) and its successor, IN-SPIRE (Hetzler et al. 2005), Topic Islands (Miller et al. 1998) and Knowledge Explorer (Novak 2007) (Figure 1).

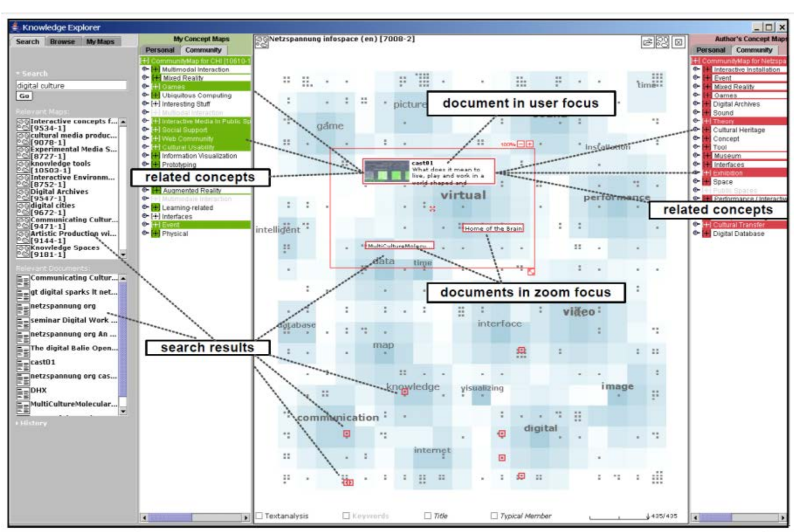

Figure 1: Knowledge Explorer (Novak 2007), a visualization that uses spatialization to create a landscape of documents.

Spatialization starts with the conversion of a particular set of documents to a multi-dimensional vector form. The number of dimensions obtained in this process is high, and some method for dimensionality reduction is usually applied to make the problem more computationally tractable. Principal Component Analysis (PCA) is one of the most common dimensionality reduction techniques (Jolliffe 2005). Others include Multi-Dimensional Scaling (MDS), Self-Organizing Maps (SOM), spring models (Skupin and Fabrikant 2003), Latent Semantic Indexing (LSI) (Fortuna, Grobelnik and Mladenić 2005), and exemplar-based spatialization (Chen et al. 2009). The process of dimensionality reduction is not trivial and the results will change depending on what technique was used. MDS, for example, can be used to produce both metric and non-metric solutions. The former preserve the pairwise distance between documents in the original vector space to the greatest possible extent, whereas the latter only preserves the *order* of the distances (Risch et al. 2008). After the number of dimensions has been reduced, the resulting documents are added to the document map. Although it is possible to plot individual documents (Crow, Pottier and Thomas 1994), clusters of documents are often mapped for enhanced readability, as large collections will cause rapid overplotting (Fortuna et al. 2005, Hetzler et al. 2005, Wise 1999).

In order to provide some meaning to the resulting visualization, the document map needs to be labeled. The decision to map either individual items or clusters of documents will also have implications for the labeling process. For individual documents, most frequent key terms can be used, whereas in case of cluster display (Figure 2), the centroid of the cluster can be used to provide a label (Risch et al. 2008). The process of label placement itself presents a number of problems. Although algorithms for automated point label placement have been studied for quite some time now (Christensen, Marks and Shieber 1995, Wagner et al. 2001), there are still debates as to the effectiveness and the quality of automatic text positioning algorithms (Van Dijk et al. 2002). Interactive labeling techniques offer additional freedom in balancing the level of detail and visual clarity (Fekete and Dufournaud 2000), whereas interactive zooming methods prompt research in the area of label generalization (Skupin 2002).

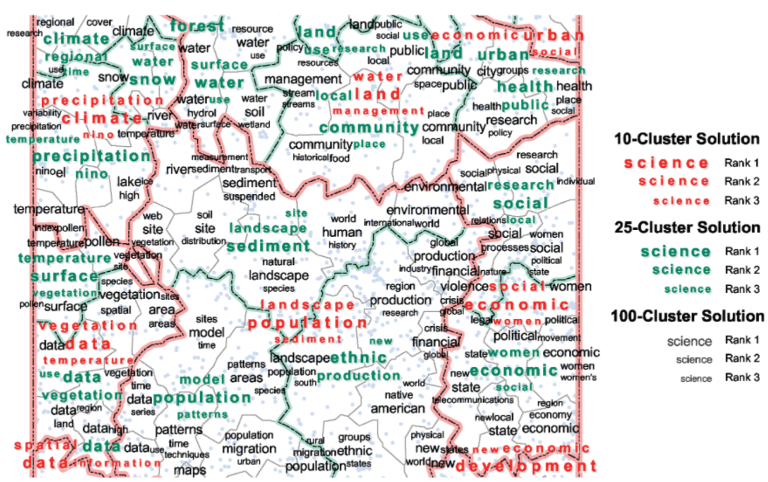

Figure 2: Skupin (2002) demonstrated techniques for labeling document clusters that use principles from cartographic design.

2.1.2 Visualization of document structure

Whereas document spatialization deals with the relationships between entire collections of documents, the *internal* structure of the document can also be looked at from the spatial perspective. More than a decade ago, Hearst (1995) introduced a display paradigm called TileBars, which mapped the content of the user query to the structure of the document in a graphical fashion. A plethora of visualization techniques similar in spirit to the original TileBars has emerged since that time, including Ink Blots (Abbasi and Chen 2007), SeeSoft (Eick, Steffen and Sumner Jr 1992), Compus (Fekete and Dufournaud 2000), as well as work by Fang et al. (2006), Keim and Oelke (2007), and Oelke et. al (2008). Some of this work (e.g. Compus) focused on exploration of document structure explicitly (Figure 3), other research (Krstajic et al. 2010) has focused on comparisons between multiple documents exclusively, whereas most provide for some combination of both (e.g. Ink Blots, SeeSoft and the original TileBars).

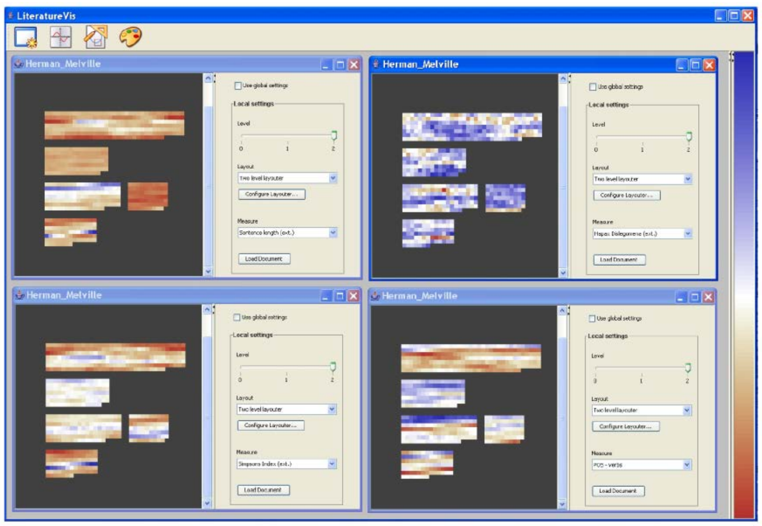

Figure 3: Keim and Oelke (2007) developed LiteratureVis to explore document structure using different computational methods. Shown here are four different methods applied to a Herman Melville's novel Moby Dick, each revealing different aspects of the novel's internal structure and composition.

Uncovering and comparing spatial references in documents has also been achieved through the use of geographic visualization techniques. Examples include SensePlace (Tomaszewski et al. 2011), a system that supports document foraging to find conceptually-similar document sets and then map their spatial "footprints" with an interactive map. The footprint visualization method in SensePlace can show the origin of a news article, and draw links outward to other placenames mentioned in an article. A similar spatial footprint visualization technique (Figure 4) was used as part of another geographic visualization tool called HealthGeoJunction focused on foraging PubMed articles about avian influenza (MacEachren et al. 2010).

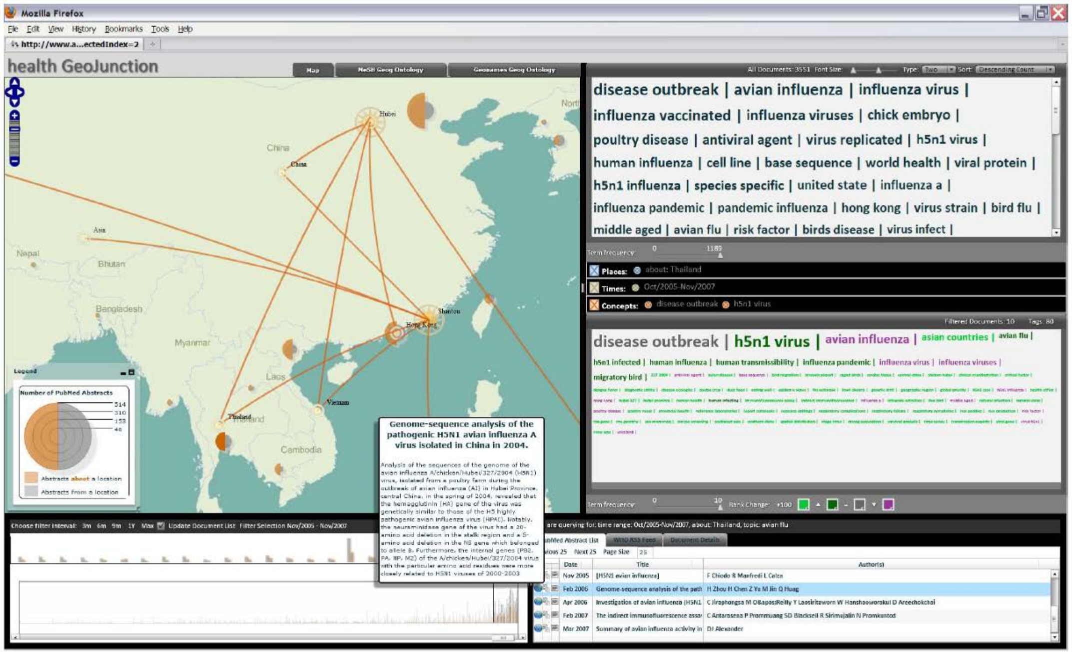

Figure 4: MacEachren et al. (2010) developed a visual technique for showing placename mentions across multiple PubMed articles on avian influenza in a tool called HealthGeoJunction. The technique uses special point symbols to highlight the origins of articles and distinguish those from other places mentioned in the article text.

2.2 Temporal characteristics of the dataset

Simultaneous display and analysis of spatial and temporal dimensions is an area of active research in geovisual analytics and the broader information visualization community. Examples of work in this domain include interactive timelines of various kinds, such as Dynamic Spiral Timeline (Chin et al. 2009) and Arc Diagrams (Wattenberg 2002). The majority of the recent papers, however, focus on iterations of two kinds of visualizations; the space-time cube and Theme River.

2.2.1 Space-Time Cube

Originally proposed by Hägerstrand (1970), and later applied in a variety of GIS-related contexts by Kraak (2003), the space-time cube metaphor uses the z-axis that is traditionally reserved for height to display time. As the subject of the analysis moves through space and time, it leaves a trace in the space-time cube. This idea has been applied to a number of domains that deal with space-time narratives (Figure 5) and is well-documented (Eccles et al. 2008, Kapler et al. 2008, Kwan 2002). One of the key limitations of the space-time cube metaphor is the low number of subjects that can be traced simultaneously. One study that sought to evaluate the utility of a space-time cube found that some simple tasks were more easily achieved using a 2D representation, while others that involved more complex analysis were easily accomplished using the space-time cube (Kristensson et al. 2009). This study, however, did not use a space-time cube approach that leveraged qualitative geographic information, so it remains to be seen if this approach has empirically-validated utility when using those data sources.

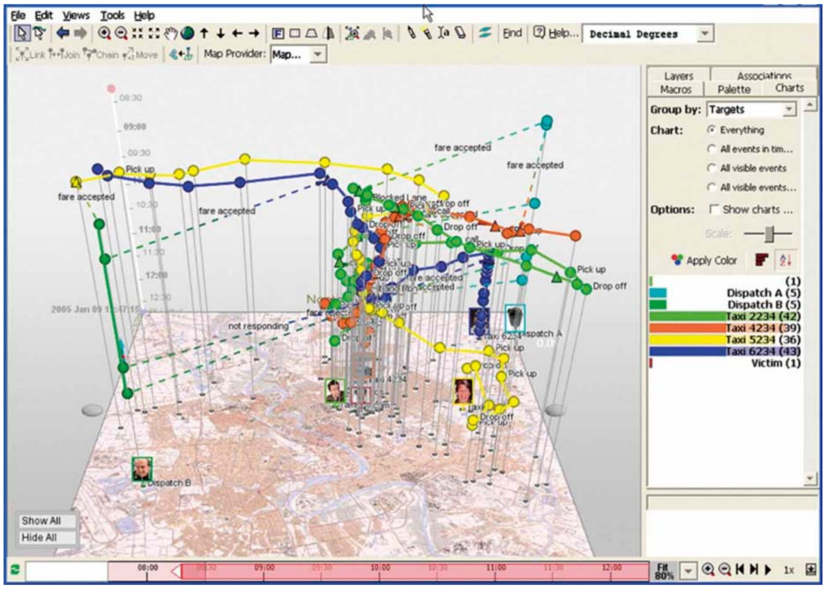

Figure 5: Eccles et al. (2008) demonstrated the use of the space-time cube metaphor for connecting spatio-temporal information to narratives to describe stages of a story, including actors and their transactions.

2.2.1 ThemeRiver

Proposed by Havre et al. (2000), the ThemeRiver representation method can show variations in thematic content over time from an associated document collection. Similar to TileBars, ThemeRiver gave birth to a large range of flow-based visualization techniques (Fang et al. 2006). One of the most successful recent iterations is included in the Visual Backchannel system by Dörk et al. (2010), which features an implementation of ThemeRiver metaphor embedded as one of its primary interactive components (Figure 6). Subsequent work by Luo et al. (2011) developed a prototype tool called EventRiver to focus explicitly on representing event data in a ThemeRiver-style representational structure. EventRiver uses computational methods to detect and define events in streams of thematic data and then represents events using a modified ThemeRiver approach. Taking the evolution of ThemeRiver further to include explicit consideration of spatial references could be envisioned as theme rivers connecting places to one another on a map, though the temporal dimension would then be conflated with the distance dimension, and scaling upward toward dozens or hundreds of possible linkages between places would result in a difficult to parse display. However, instead of using ThemeRiver's implicit focus on the temporal dimension, one could envision thematic information as overlays on maps as has been done for decades using dasymetric mapping techniques (Slocum et al. 2005), much like what has been demonstrated already in topic landscape research (Skupin and Fabrikant 2003).

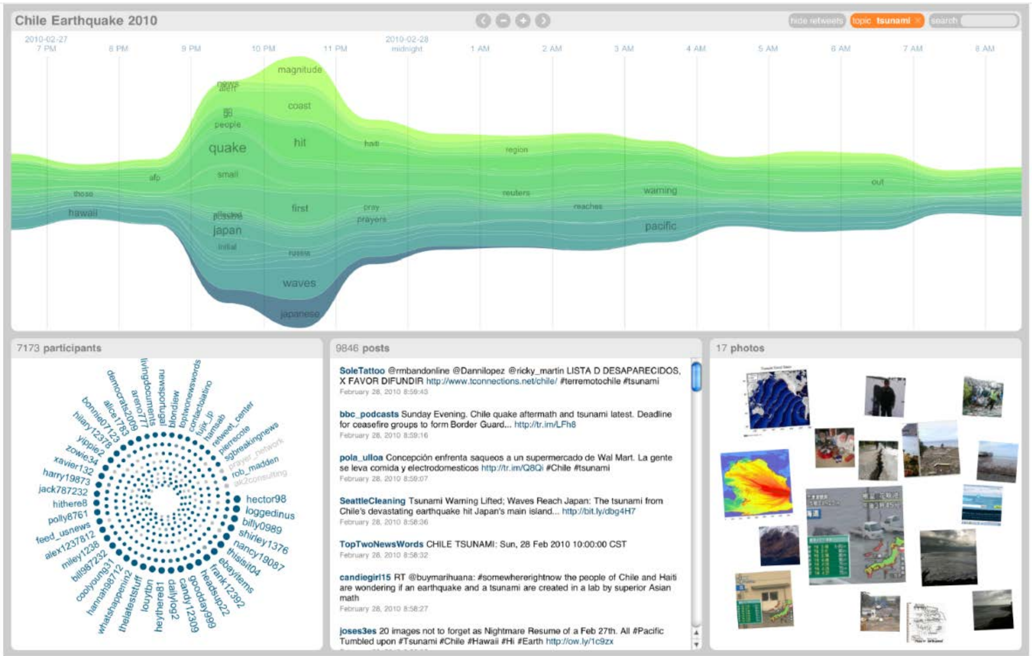

Figure 6: Dörk et al. (2010) used a ThemeRiver display to represent thematic information emerging from streaming Twitter data in the wake of the recent Chile Earthquake. This visualization technique reveals changes in what people were talking about prior to, during, and after the disaster.

2.3 Spatial characteristics of the dataset

Whereas the spatialization techniques described above deal with the notion of space as a metaphor, geovisual analytics is concerned with data that is explicitly geographic. Geographical information can be stored as part of the original metadata (e.g. GPS coordinates describing the location of the tweet), it can be discovered through the process of entity extraction (e.g. textual references to a populated place, written directions, etc.), or they can arise as a product of data processing (e.g. area of influence for a particular newspaper) (Mehler et al. 2006). The spatial nature of the geographic information will define the variables available to the geovisual analyst. Individual point records are easy to map, yet they tend to cause overplotting and can be difficult to analyze, particularly when the number of records is high. Certain phenomena are point features in nature, but cannot be positioned exactly due to uncertainty in the spatial information available. Features that have a certain area to them (populated places, parks, historical districts and neighborhoods, etc.) may be easier to map due to the existence of physical or cultural boundaries, but they too are often imperfectly defined and cannot be precisely positioned. Spatial aggregates of geographic data based on a regular grid can provide access to summary statistics for the underlying dataset and are easy to obtain, but they suffer from the Modifiable Areal Unit Problem (MAUP) (Gehlke and Biehl 1934) and require resampling to multiple resolutions for more advanced exploration. Regardless of the nature of available geographic information available, two approaches towards mapping text can be identified: using coordinated views to connect text to symbols on a map, and mapping text as an overlay by itself.

2.3.1 Coordinated-view visualizations

The idea behind the concept of coordinated-view, spatially-anchored visualizations is to use multiple views to connect text representations to maps and support cross-filtering between them. Examples include mapping a tag cloud that collects data from a particular location (Slingsby et al. 2007) and plotting tree maps over a regular spatial grid (Slingsby, Dykes and Wood 2008). Although this approach allows for reuse of existing techniques, it can come at the expense of screen space (Figure 7).

Figure 7: Slingsby et al. (2007) used linked views to pair an interactive map with an interactive tag cloud. Users can click on the tag cloud to filter the map and zoom to an area of interest where tags are located.

2.3.2 Mapping text as an overlay

There have been a number of attempts to represent the textual information directly onto the geographic scale in question. Based on the concept of relevancy, Jaffe et al. (2006) maps a selection of tags describing the geographical region that is currently in view. World Explorer (Ahern et al. 2007) uses the criteria of spatial locality to identify the most salient content. Tags that occur in a concentrated area and are sparse in its neighborhood are deemed to be more representative of that area, and receive a higher rank (Figure 8). Both techniques modify the selection of tags based on the current zoom level, aiming to preserve context and provide detail upon demand, a concept identical to cartographic generalization (Skupin 2002) that has been labeled *semantic zoom* (Jaffe et al. 2006).

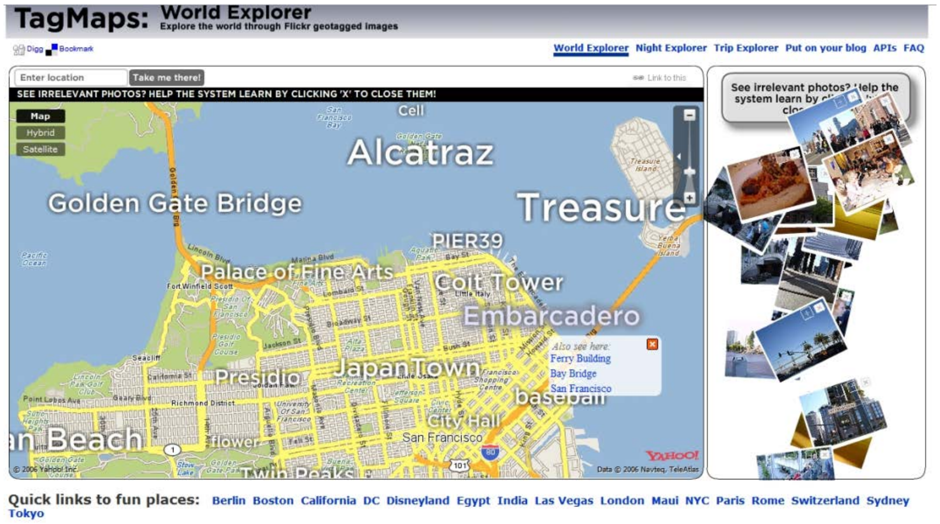

Figure 8: World Explorer, developed by Ahern et al. (2007), anchors tags directly to their spatial locations where they were mentioned or observed.

2.4 Spatial characteristics of text as a design element

Despite certain successes, the majority of the studies we reviewed ignore some of the fundamental properties of text as a visual artifact that includes spatial dimensions that have meaning in a geographical context. These properties must be addressed in the framework of geovisual analytics.

2.4.1 Position of text as a proxy for relevance

TagOrbitals (Kerr 2006) is a recent modification of tag cloud that uses the metaphor of an orbit to show relevance. Relevant tags are positioned in the inner orbits, with everything else moved out to the periphery (Figure 9). This idea can be extended to the visualization of text in a geographical context. With the center of the cluster anchored in geographical space, elements can be deemed to be more or less representative of a particular location based on their position.

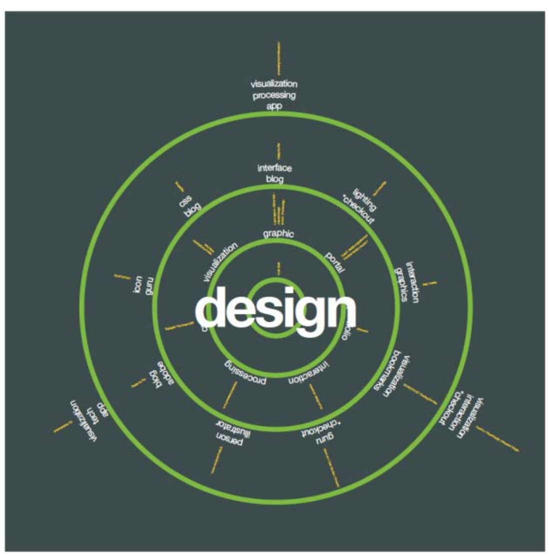

Figure 9: TagOrbitals (Kerr 2006) could be modified to incorporate direction/distance for geographic data.

2.4.2 Text footprint as a proxy for geographical character of phenomena described
Much criticism of current text visualization techniques stems from the fact that there is often a mismatch between the footprint of the label and the footprint of the phenomenon it describes. In traditional cartography, this is resolved through careful manipulation of label positioning and the font face to ensure that the ambiguity is minimal. This cartographic expertise can be applied to the automated visualization of text as well through the application of placement rules, in a manner similar to what has been achieved by Esri with their ArcGIS Maplex label placement tools.

2.4.3 Text overlap as a proxy for geographical interaction
Text overlap is most often seen as a problem. Indeed, with the weight of the textual items being roughly similar, any amount of overlap can render one (or more) labels unreadable. Creative use of overlap in combination with different font faces, however, can be used to emphasize the nature of the conversation (regional trend versus localized topic, for example), or a limited amount of temporal information (past and current topics). This would allow visualization designers to preserve the balance between context and detail, an important goal we have previously identified.

2.5 Considerations for interface design
As a result of our review, a number of persistent common design themes have been identified. Some of these themes revolve around successful metaphors for user interface design, others put emphasis on the integration of multiple components into the working system. We classify these themes into two categories; key visualization principles, and potential modes of interaction. Below we synthesize examples for each category.

2.5.1 Key visualization principles
1. *Distance as a metaphor for similarity* (Fabrikant et al. 2004, Wise 1999). In the majority of the examples we have reviewed, distance between entities is commonly used as a measure of their similarity. Distance can be measured in linear fashion or across the network, and it is subject to modification by other visual variables, yet it persists as one of the most powerful tools in the designer's arsenal. This design guideline corresponds closely to what many regard as a fundamental principle in the science of Geography, Tobler's first law of Geography (Tobler 1970) , which states, "Everything is related to everything else, but near things are more related than distant things."
2. *Generalization and semantic zoom* (Jaffe et al. 2006, Skupin 2002). Proper generalization and pruning of textual information is one of the most important factors for a successful design. Approaches to generalization differ from intelligent elimination of extra labels to the development of semantic summaries for an underlying dataset.
3. *Ordering and similarity* (Fekete and Dufournaud 2000, Hassan-Montero and Herrero-Solana 2006, Keim and Oelke 2007, Oelke et al. 2008, Viégas 2005). The notion of visual similarity, visual patterns, and even visual "fingerprints" is quite prominent in the majority of the work we reviewed that deals with document structure visualization. A promising direction for future work seems to be in automated ordering and matching of such patterns.
4. *Intelligent highlighting of text* (Abbasi and Chen 2007, Miller et al. 1998, Obendorf and Weinreich 2003). Despite all attempts to reduce the amount of text that is meant to be read,

some of the information cannot be reduced much compared to its original form. In order to facilitate rapid knowledge collection, strategies for intelligent highlighting of the textual information are investigated. Such highlights are used to identify changes in topic, unusual names and places, and other important features in the document. A variety of visual methods have been proposed to extend simple color-based highlighting in geovisualization systems (Robinson 2009), and these too may be worth exploring in relation to interactive environments that make use of geospatially-oriented qualitative information.

5. *User-centered visualization* (Ahn and Brusilovsky 2009, Elmqvist and Tsigas 2007, Novak 2007). As part of learning and collaborative effort, a number of studies focused on capturing the expertise and interests of a user in order to structure the form of the visualizations more effectively. It is possible to imagine new visualization metaphors to support different portions of the analytical process, including specialized views designed to support *exploration* or *synthesis* (Robinson 2011)..

2.5.2 Potential modes of interaction

A number of possible interaction techniques that are particularly fit for the exploration of multi-dimensional textual data have been identified:

1. *Focus and context zooming* techniques were described by Qu (2009) as a flexible way to construct navigation paths in urban environment. Although originally designed for a different purpose, this visualization technique seems well-fit for any environment that has high amounts of mutual occlusion between its features, such as maps that include large collections of georeferenced text. A key aspect of the technique described by Qu (2009) is that it can be automatically applied to highlight items of interest that match a particular query.
2. *Radial representation* methods like Sunburst by Stasko et al. (2007) and TagOrbitals by Kerr (2006) can include modifications of the treemap concept that use a radial layout. Creative use of the space on the outside of a circular treemap allows for seamless dynamic exploration of smaller categories. Because radial representations can incorporate concepts related to cardinal direction and distance from the center, it may be possible to order and visualize text using such methods while incorporating some key aspects of geographic space.
3. *Lens distortion* techniques like Balloon focus by Tu and Han-Wei (2008) solve the problems associated with zooming in on a treemap without requiring changes to shape or topology. Existing categories are dynamically rescaled to make space for the ones in focus. Methods like this one and others that do manipulate shape and topology to create so-called fisheye views (Robertson and Mackinlay 1993) may be suitable as user-driven interactive means for exploring dense geographic information landscapes using a mouse or other input device.
4. *Coordinated multiple-view* approaches (Roberts 2007) are the most common structure for multi-representational visualization tools, and they constitute a key method for integrating traditional map-based visualizations with textual analysis components. Coordinated multiple-view systems use synchronized interactions and visual cues to support queries across representational forms. Such approaches forgo the goal of developing a single view that incorporates all forms and

interactions in favor of an ecosystem of multiple representational views that are closely intertwined to support dynamic work.

3. New Approaches for Representing Qualitative Geographic Information with Text

Our review of previous work has highlighted some common themes in existing approaches for representing and interacting with text, and suggests some fruitful ideas for future methods to pursue in support of a future in which qualitative geographic information is more readily explored and evaluated in the context of geospatial analysis. As part of this review, we note that little has been done so far to design text representation techniques that directly tie to common characteristics of geographic data that would support analysis using overlays on a map. Current approaches simply place tags over geographic space, while the design of the tag itself has had relatively little attention to its visual design potential. Here we propose some initial design directions to advance the state of the art.

The design concepts shown in the following sections reveal how thematic, temporal, sentiment, and certainty information associated with qualitative geographic information can be represented by manipulating map labels alone. Typefaces, their attributes (bold weighting, italics, underlines, etc…), colors, textures, shadowing, and other visual effects can be combined to highlight a range of values for thematic, temporal, sentiment, and certainty information. To our knowledge, there has been little prior work focused on specifying such methods to support analysis of qualitative geographic information. The methods we propose here could be created dynamically, but are designed to work in static presentations to ensure maximum dissemination potential as maps are shared outside of the interactive environments in which they are generated.

Each design concept shows map labels that are sized according to their frequency from a simulated social media dataset that has been explored using computational methods to identify references to geographic locations. In the examples shown below, Pennsylvania's state label is large, indicating high frequency (many mentions) in this simulated dataset, while the city of Erie is smaller because it appears less frequently (fewer mentions). Each label is styled using the scale shown at the top of the mockup. The styles indicate each location's membership in a second class of categories (thematic, temporal, sentiment, or certainty). For example, Pennsylvania may have been mentioned many times, but these mentions may have occurred mostly at the beginning of the available time range, so the label would receive the corresponding "old" look/feel (Figure 11). Styles are assigned at random to the map to simulate the variability that one would expect from a real dataset.

3.1 Thematic Information

Many sources of geographically-oriented text include categories and other thematic information that can provide useful context. To represent categories using text alone, a range of visual variables are potentially applicable. Sequential categories can be displayed by manipulating basic aspects of a single typeface, as shown in Figure 10. Here, a single typeface is used to represent a particular data stream, while changes to this typeface are applied to create a sequential visual effect to introduce the notion of categories in a sequence from low to high.

Figure 10: Sequential categorization applied to geographically-oriented text.

Text can also be manipulated to represent non-sequential categories, as shown in Figure 11. Here, typefaces are varied across the categories to include modern and classic variations on serif and non-serif forms, as well as scripted forms. The encoding is then reinforced through the use of a qualitative color scheme (using a 5-class scheme from colorbrewer.org).

Figure 11: Qualitative categories can be assigned to location mentions on a map as shown here.

Further variations are also possible for displaying thematic information through the manipulation of the text itself. Different textures could be applied to indicate multiple unique categories (hashes, dots, solid lines, etc…) or a sequence of categories (smooth to rough, for example).

3.2 Time

In addition to thematic information, many text sources include information regarding the time they were generated or last shared. Temporal representation can be encoded through changes in the typeface itself, the typeface color, and through the application of textures and shadows. An overall effect can be developed through combinations of these variables to show text that looks old or new. It's also possible to envision predicted items shown with a visual aesthetic that conveys a notion of the future. Figure 12 shows an example of how a range of visual variables and typeface controls can be combined to create a temporal scale.

Figure 12: Temporal information, from old, to new, to the future can be revealed through label design, as shown here.

3.3 Sentiment

Qualitative geographic data generated by human narratives can often reveal how people feel about a particular topic. A wide range of methods have been developed in recent years to gauge human sentiment from text, and sentiment measures can be represented on maps through graphic changes to typefaces. Figure 13 shows an example where color, typefaces, and typeface attributes are combined to reveal a 5-point negative/positive scale to show sentiment. Italics are used for negative attitudes, modern-serif fonts in medium tones are used for neutral attitudes, and modern sans-serif fonts are used to indicate positive attitudes.

Figure 13: Sentiment categorization can be visualized through changes to typeface design attributes.

3.4 Certainty

Geographic information of all kinds includes various aspects of uncertainty, either implicit due to the methods by which it was created or explicit by virtue of its timeliness, resolution, or spatial coverage. Qualitative geographic information poses special challenges, as it is often generated by individuals, and some individuals are more credible sources than others. This challenge also provides an opportunity for representation of this information on maps. Figure 14 shows an example of how a single typeface can be manipulated using color, texture, shadowing, and weight to indicate a range from least to most certain.

Figure 14: Certainty can be shown on text labels by manipulating text color, texture, shadowing, and weight.

3.5 Dynamic Methods

While the methods proposed above provide first steps toward implementable techniques that could be used across a wide range of geographic information systems that feature qualitative geospatial information, they are deliberately designed to function as static entities to support information sharing and dissemination. In most analytical scenarios, interactive tools are used to generate static captures that are then sent onward to stakeholders and decision makers.

In the future, we can expect better interactive systems that support the dissemination of interactive product to help teams of collaborators and decision makers in complex tasks. Such systems will provide the opportunity for us to develop dynamic methods that may use animation techniques (both stepwise through several course stages and smoothed, continuous animations) to reveal attributes related to themes, temporal information, sentiment, and certainty.

Such dynamic methods can be driven by explicit interaction; for example the movement of a cursor or other input device over a particular label could cause it to animate to reveal something about its associated certainty or sentiment. Dynamic methods could also be triggered by streaming information that fits a particular set of criteria; for example, a microblogging feed may be monitored to highlight information that matches common disaster type keywords, and locations on the map may change their form dynamically to show that they are becoming associated with those keyword categories.

4. Next Steps and Recommendations

Our review of existing approaches and design proposals for several new techniques provides a broad base of competing technologies and design strategies to choose from for future research and development. Our charge for the remainder of the project period is to implement some of these techniques and work toward an evaluation that can reveal whether or not some approaches are better than others. A key target going forward is the development of enough empirical knowledge to be able to rank and evaluate a wide range of methods for their suitability across a range of common analytical tasks. Such knowledge could feed into a framework much like one previously developed by MacEachren (1994) for highlighting the relative suitability of visual variables for symbolizing different data types. Out of the large number of sample techniques we reviewed here, only a handful include any results from user evaluations, and to our knowledge nobody has attempted such a synthesis so far to rank and arrange techniques according to their suitability/utility for different types of text visualization applications. While it may be impractical to develop a complete version of such a framework in the remainder of our current project period, below we suggest some next steps in this effort that would pave the way toward such an outcome.

Specifically, we recommend the following path forward for new text visualization methods and interaction techniques to be incorporated into SensePlace2, our test bed visualization environment:

1. Visualization goals:
    a. Support the use of text styling to indicate key aspects of geographic data as depicted in Section 3 (thematic, temporal, sentiment, and certainty information)
    b. Develop a radial visualization method that builds on prior work and focuses on incorporating the geographic dimension more explicitly through directional angles and distance
    c. Design dasymetric mapping methods that can show dominant themes as overlays on the map

2. Interaction goals:
    a. Implement dynamic focusing using focus+context and/or lens techniques
    b. Use coordinated multiple-view approach to connect new visualization techniques (e.g. a radial display as proposed above) to existing interactive systems

Accomplishing these goals would constitute a large step forward toward the goal of supporting analysts with better geospatial systems that integrate qualitative geographic information. While there remain many additional research and application opportunities aside from those we recommend pursuing here, we believe those we have highlighted here and evaluating them with end-users will reveal which visualization and interaction paradigms are the most valuable to pursue for next-generation systems that incorporate qualitative geographic information.

5. References

Abbasi, A. & H. Chen. 2007. Categorization and analysis of text in computer mediated communication archives using visualization. In *7th ACM/IEEE-CS Joint Conference on Digital libraries*, 11-18. Vancouver, BC.

Ahern, S., M. Naaman, R. Nair & J. Yang. 2007. World Explorer: Visualizing Aggregate Data from Unstructured Text in Geo-Referenced Collections. In *7th ACM/IEEE-CS Joint Conference on Digital Libraries*, 1-10. Vancouver, BC.

Ahn, J. & P. Brusilovsky (2009) Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization,* 8**,** 167-167.

Andrienko, G., N. Andrienko, P. Jankowski, D. Keim, M. J. Kraak, A. M. MacEachren & S. Wrobel (2007) Geovisual analytics for spatial decision support: setting the research agenda. *International Journal of Geographical Information Science,* 21**,** 839-857.

Chen, Y., L. Wang, M. Dong & J. Hua (2009) Exemplar-based Visualization of Large Document Corpus. *IEEE Transaction on Visualization & Computer Graphics,* 15**,** 1161-1168.

Chin, G., M. Singhal, G. Nakamura, V. Gurumoorthi & N. Freeman-cadoret (2009) Visual analysis of dynamic data streams. *Information Visualization,* 8**,** 212-212.

Christensen, J., J. Marks & S. Shieber (1995) An empirical study of algorithms for point-feature label placement. *ACM Transactions on Graphics (TOG),* 14**,** 203-232.

Cidell, J. (2010) Content clouds as exploratory qualitative data analysis. *Area*.

Crow, V., M. Pottier & J. Thomas. 1994. Multidimensional visualization and browsing for intelligence analysis. Pacific Northwest Lab., Richland, WA (United States).

Dörk, M., D. Gruen, C. Williamson & S. Carpendale (2010) A Visual Backchannel for Large-Scale Events. *IEEE Transaction on Visualization & Computer Graphics,* 16**,** 1129-1138.

Eccles, R., T. Kapler, R. Harper & W. Wright (2008) Stories in GeoTime. *Information Visualization,* 7**,** 3-17.

Eick, S. C., J. L. Steffen & E. E. Sumner Jr (1992) Seesoft: a tool for visualizing line oriented software statistics. *IEEE Transactions on Software Engineering,* 18**,** 957-968.

Elmqvist, N. & P. Tsigas (2007) CiteWiz: a tool for the visualization of scientific citation networks. *Information Visualization,* 6**,** 215-232.

Fabrikant, S. I., D. R. Montello, M. Ruocco & R. S. Middleton (2004) The Distance–Similarity Metaphor in Network-Display Spatializations. *Cartography and Geographic Information Science,* 31**,** 237-252.

Fang, S., M. Lwin & P. Ebright. 2006. Visualization of unstructured text sequences of nursing narratives. In *ACM Symposium on Applied Computing*. Dijon, France.

Fekete, J. D. & N. Dufournaud. 2000. Compus: visualization and analysis of structured documents for understanding social life in the 16th century. In *5th ACM International Conference on Digital Libraries*, 47-55. San Antonio, TX: ACM.

Fortuna, B., M. Grobelnik & D. Mladenić (2005) Visualization of Text Document Corpus. *Informatica,* 29**,** 497-502.

Gehlke, C. & H. Biehl (1934) Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association,* 29**,** 169-170.

Hägerstrand, T. (1970) What about people in Regional Science? *Papers in Regional Science,* 24**,** 6-21.

Hassan-Montero, Y. & V. Herrero-Solana. 2006. Improving tag-clouds as visual information retrieval interfaces. In *International Conference on Multidisciplinary Information Sciences and Technologies*, 1-6. Merida, Spain.

Havre, S., B. Hetzler & L. Nowell. 2000. ThemeRiver: Visualizing Theme Changes over Time. In *IEEE Symposium on Information Visualization*, 115-123. Salt Lake City, UT: IEEE.

Hearst, M. A. 1995. TileBars: visualization of term distribution information in full text information access. In *ACM Conference on Human Factors in Computing Systems (CHI '95)*, 59-66. Denver, CO.

Hetzler, E. G., V. L. Crow, D. A. Payne & A. I. Turner. 2005. Turning the Bucket of Text into a Pipe. In *IEEE Symposium on Information Visualization,* eds. J. Stasko & M. Ward, 89-94. Minneapolis, MN: IEEE Computer Society.

Jaffe, A., M. Naaman, T. Tassa & M. Davis. 2006. Generating summaries and visualization for large collections of geo-referenced photographs. In *8th ACM International Workshop on Multimedia Information Retrieval (MIR '06)*, 89-98. Santa Barbara, CA.

Jolliffe, I. T. 2005. *Principal Component Analysis*. New York: Wiley.

Kapler, T., R. Eccles, R. Harper & W. Wright. 2008. Configurable Spaces: Temporal Analysis in Diagrammatic Contexts. In *IEEE Symposium on Visual Analytics Science and Technology*, 43-50. Columbus, OH.

Keim, D. A. & D. Oelke. 2007. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *IEEE Symposium on Visual Analytics Science and Technology*, 115-122. Sacramento, CA.

Kerr, B. 2006. TagOrbitals: a tag index visualization. In *ACM Special Interest Group on Computer Graphics and Interactive Techniques Conference (SIGGRAPH '06)*. Boston, MA.

Kraak, M. J. 2003. The space-time cube revisited from a geovisualization perspective. In *21st International Cartographic Conference (ICC 2003)*. Durban, South Africa.

Kristensson, P. O., N. Dahlback, D. Anundi, M. Bjornstad, H. Gillberg, J. Haraldsson, I. Martensson, M. Nordvall & J. Stahl (2009) An evaluation of space time cube representaion of spatiotemporal patterns. *IEEE Transaction on Visualization & Computer Graphics,* 15**,** 696-702.

Krstajic, M., E. Bertini, F. Mansmann & D. A. Keim. 2010. Visual analysis of news streams with article threads. In *First International Workshop on Novel Data Stream Pattern Mining Techniques (StreamKDD '10)*, 39-46. Washington, DC.

Kwan, M.-P. (2002) Feminist Visualization: Re-envisioning GIS as a Method in Feminist Geographic Research. *Annals of the Association of American Geographers,* 92**,** 645-661.

Lee, B., N. H. Riche, A. K. Karlson & S. Carpendale (2010) SparkClouds: visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics,* 16**,** 1182-1189.

Luo, D., J. Yang, M. Krstajic, W. Ribarsky & D. Keim (2011) EventRiver: Visually exploring text collections with temporal references. *IEEE Transactions on Visualization and Computer Graphics,* To appear.

MacEachren, A. M. 1994. *SOME Truth with Maps: A Primer on Design and Symbolization*. Washington, D. C.: Assocation of American Geographers.

MacEachren, A. M. & M. J. Kraak (1997) Exploratory cartographic visualization: advancing the agenda. *Computers & Geosciences,* 23**,** 335-343.

MacEachren, A. M., M. S. Stryker, I. J. Turton & S. Pezanowski (2010) HEALTH GeoJunction: place-time concept browsing of health publications. *International Journal of Health Geographics,* 9.

Mehler, A., B. Yunfan, L. Xin, W. Yue & S. Skiena (2006) Spatial Analysis of News Sources. *IEEE Transactions on Visualization and Computer Graphics,* 12**,** 765-772.

Miller, N., P. Wong, M. Brewster & F. H. 1998. Topic Islands - A wavelet-based text visualization system. In *IEEE Symposium on Information Visualization*, 189-196. Research Triangle Park, N.C.

Novak, J. 2007. Helping Knowledge Cross Boundaries: Using Knowledge Visualization to Support Cross-Community Sensemaking. In *40th Annual Hawaii International Conference on System Sciences (HICSS 2007)*, 38-38. Waikoloa, HI.

Obendorf, H. & H. Weinreich. 2003. Comparing Link Marker Visualization Techniques: Changes in reading behavior. In *12th International World Wide Web Conference*, 736-745. Budapest, Hungary.

Oelke, D., P. Bak, D. A. Keim, M. Last & G. Danon. 2008. Visual evaluation of text features for document summarization and analysis. In *IEEE Symposium on Visual Analytics Science and Technology*, 75-82. Columbus, OH.

Qu, H., W. Haomian, C. Weiwei, W. Yingcai & C. Ming-Yuen (2009) Focus+Context Route Zooming and Information Overlay in 3D Urban Environments. *IEEE Transactions on Visualization and Computer Graphics,* 15**,** 1547-1554.

Risch, J., A. Kao, S. Poteet & Y. Wu. 2008. Text Visualization for Visual Text Analytics. In *Visual Data Mining*, 154-171.

Roberts, J. C. 2007. State of the art: coordinated & multiple views in exploratory visualization. In *Fifth International Conference on Coordinated and Multiple Views in Exploratory Visualization*, 61-71. Zurich, Switzerland: IEEE Computer Society.

Robertson, G. G. & J. D. Mackinlay. 1993. The document lens. In *6th Annual ACM Symposium on User Interface Software and Technology*, 101-108. San Jose, CA.

Robinson, A. C. 2009. Visual highlighting methods for geovisualization. In *24th International Cartographic Conference*. Santiago, Chile.

Robinson, A. C. (2011) Supporting Synthesis in Geovisualization. *International Journal of Geographical Information Science,* 25**,** 211-227.

Sinclair, J. & M. Cardew-Hall (2008) The folksonomy tag cloud: when is it useful? *Journal of Information Science,* 34**,** 15-15.

Skupin, A. (2002) A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications,* 22**,** 50-58.

Skupin, A. & S. I. Fabrikant (2003) Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science,* 30**,** 99-119.

Slingsby, A., J. Dykes & J. Wood (2008) Using treemaps for variable selection in spatio-temporal visualisation. *Information Visualization,* 7**,** 210-210.

Slingsby, A., J. Dykes, J. Wood & K. Clarke. 2007. Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets. In *IEEE International Conference on Information Visualization*, 497-504. Zurich, Switzerland.

Slocum, T., R. B. McMaster, F. C. Kessler & H. H. Howard. 2005. *Thematic Cartography and Geographic Visualization*. Upper Saddle River, NJ: Pearson Education.

Stasko, J., C. Gorg, L. Zhicheng & K. Singhal. 2007. Jigsaw: supporting investigative analysis through interactive visualization. In *IEEE Symposium on Visual Analytics Science and Technology (VAST 2007)*, 131-138. Sacramento, CA.

Thomas, J. J. & K. A. Cook. 2005. Illuminating the path: the research and development agenda for visual analytics. New York: IEEE CS Press.

Tobler, W. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography,* 46**,** 234-240.

Tomaszewski, B., J. Blanford, K. Ross, S. Pezanowski & A. M. MacEachren (2011) Supporting geographically-aware web document foraging and sensemaking. *Computers, Environment, and Urban Systems,* 35**,** 192-207.

Tu, Y. & S. Han-Wei (2008) Balloon Focus: a Seamless Multi-Focus+Context Method for Treemaps. *IEEE Transactions on Visualization and Computer Graphics,* 14**,** 1157-1164.

Van Dijk, S., M. Van Kreveld, T. Strijk & A. Wolff (2002) Towards an evaluation of quality for names placement methods. *International Journal of Geographical Information Science,* 16**,** 641-662.

Viégas, F. B. 2005. Revealing individual and collective pasts: Visualizations of online social archives. 126. Boston: Massachusetts Institute of Technology, Media Arts and Sciences.

Viégas, F. B., M. Wattenberg & J. Feinberg (2009) Participatory Visualization with Wordle. *IEEE Transaction on Visualization & Computer Graphics,* 15**,** 1137-1144.

Wagner, F., A. Wolff, V. Kapoor & T. Strijk (2001) Three rules suffice for good label placement. *Algorithmica,* 30**,** 334-349.

Wattenberg, M. 2002. Arc diagrams: Visualizing structure in strings. In *IEEE Symposium on Information Visualization*, 110-116. Boston, MA.

Wise, J. A. (1999) The ecological approach to text visualization. *Journal of the American Society for Information Science,* 50**,** 1224-1233.

Wood, J., J. Dykes, A. Slingsby & K. Clarke (2007) Interactive visual exploration of a large spatio-temporal datasaet: reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics,* 13**,** 1176-1183.